Economics and Sociology
Occasional Paper No. 291

A NOTE ON DATA COLLECTION AND ANALYSIS PROBLEMS IN RESEARCH ON RURAL FINANCIAL MARKETS

by

Richard L. Meyer
Agricultural Economist

Department of Agricultural Economics and Rural Sociology
The Ohio State University
2120 Fyffe Road
Columbus, Ohio 43210

A NOTE ON DATA COLLECTION AND ANALYSIS PROBLEMS IN RESEARCH ON RURAL FINANCIAL MARKETS*

by
Richard L. Meyer
The Ohio State University

I. Introduction

In another paper prepared for this workshop, Dale Adams called attention to the expansion in research on rural financial market problems which has occurred in many low income countries (LIC) in recent years. In this note, I would like to comment on some of the problems which have arisen, particularly with respect to farm level studies, as policy related research has been expanded. A number of my observations are conditioned by our experience at The Ohio State University with the recently completed Capital Formation Project funded by USAID. A significant portion of the research dealt with the relationship between agricultural credit and farm growth.

In this short paper, I would like to focus on three issues related to research on rural financial markets in LIC's: 1) the comparative advantages and interests of various types of researchers need to be more carefully considered when allocating opportunities and responsibilities for research on rural financial markets, 2) the several existing sources of primary data must be more carefully identified and evaluated before mounting huge new farm level surveys, and 3) the relatively ignored practical problems of collecting and analyzing large quantities of primary data require greater attention when planning and funding research projects.

---

* Paper prepared for the Workshop on Rural Financial Markets, Bangkok, Thailand, October 9-11, 1975.

II. The Division of Labor in Research

Several different types of individuals and institutions comprise the research community interested in and capable of conducting research on rural financial markets in any country. This community includes, on the one hand, isolated individuals with various teaching, research and administrative duties, and on the other hand, well-organized research and administrative institutes and offices frequently found in central and commercial banks, state and federal government offices, universities, and private consulting firms. Each of these individuals and groups have particular advantages, competencies and limitations with respect to certain types of research. While they may be especially suited to research one specific subject, they may be quite inappropriate for another. Some can devote considerable time to short-term projects with quick payoff, while others are best suited to long-term research. For universities, the learning involved in research conducted by students may be more valuable than the results themselves.

The demand for research likewise comes from several individuals and groups ranging from governmental policymakers and bank directors to associations of farmers and input supply firms.

Unfortunately this diversity of competencies and needs has not been clearly recognized and the result has often been a mismatch or incompatibility between the supplier and user of research results. All too frequently universities and academically oriented researchers have been expected to do "fire fighting" where the answer to a specific crucial problem was needed yesterday by a policymaker. On other occasions, these same researchers have been prevented from publishing the results of their

work because of restrictions set by the contractor of the research. This failure to recognize the researchers' interest in careful long-term studies, methodological advances, and publishing in prestigious journals frustrates both the supplier and user of research.

What is required is a division of labor. Policymakers need to develop analytical capabilities, perhaps in their respective institutions, which can address the legitimate immediate questions requiring quick, pragmatic but well-reasoned judgments regarding policies. Many decisions simply cannot wait for the results of long-term research. Simultaneously, research linkages must be built with individuals and institutions capable of doing the longer-term, more sophisticated research and supplying a continuous flow of information and ideas about policies, the benefit of which is captured by the society as a whole perhaps long after a particular policymaker has come and gone.

Related to this general issue of compatibility is the question of adequately integrating researchers into the policymaking system so that their output is timely and relevant. Policymakers justifiably complain that much academic research has no immediate applicability to their specific problems but fail to recognize that this is due in part because researchers have often been marginalized from the whole policy-making process. Only through continual interaction with the policy-makers can researchers hope to produce that information which will make an immediate significant contribution to pressing problems.

III. Data Sources

When farm level research on rural financial markets is proposed in many LIC's, the first reaction is frequently to organize a cross-sectional survey of farmers. This may be the only alternative if the

question to be studied is very unique and specific. But in many countries an amazing amount of data have already been collected through special household surveys or for census purposes, and valuable insights could be obtained on general financial issues if the data would be analyzed for this purpose. Scarce human and financial resources dictate that these sources be carefully identified and evaluated before mounting new expensive, time-consuming surveys.

Financial institutions also contain a wealth of information, much of which has not been used for research. Unfortunately, it is usually not stored and collated in a very accessible form. Many banks, for example, frequently collect important financial and economic data each year when borrowers take out operating loans, but the data are archived (or worse yet destroyed) when the loan is repaid rather than held for the following year in the individual farmer's file. Thus any researcher attempting to study the progress of farmers through time is stuck with an almost impossible task of locating a specific file for each of the years that the farmers did business with the firm.

Simplifying administrative procedures would go a long way toward reducing the quantity of records required on any individual and facilitate the filing of data for several years in the individual's credit folder. In addition, high priority must be given to carrying over basic information from one year to the next. Obviously these data are limited: questions dealing with all farmers cannot be addressed by studying only those that do business with a particular institution. But at least the role and problems of one institution would become clearer by analyzing the history of some of its clients over time rather than concentrating on the characteristics of a particular group that does business with it in

any one year.

An interesting tendency seems to prevail in those cases where conducting a survey is the only alternative for data collection. Policy-makers and researchers alike appear convinced that thousands of observations are necessary for reliable research. Granted, the heterogeneity of agriculture requires that data be collected from a vast number of sizes and types of farms located in a variety of regions if we are to completely understand certain issues for the entire sector. But the time and resources needed for generating and analyzing such information simply requires that we content ourselves with a more partial approach. We must hope to capture the main features of the question studied through less ambitious surveys, for reasons that will be explained a bit later, even if the statistical representativeness of the observations implies caution in generalizing to the entire farm population. In other words, we must recognize the tradeoff between representativeness and time and resources.

Another bias seems to be evident in many countries regarding farm level data. Cross-sectional surveys are frequently conducted at great expense, while little effort is placed on developing longitudinal data from a selected panel of farms. Even traditional agriculture experiences change, so the validity of information collected in one cross-section deteriorates over time. Furthermore, some issues can only be adequately studied with time series data. The implication is that the efforts of a few countries must be imitated elsewhere: some of the resources used on periodic surveys must go into beginning the periodic collection of information from a selected panel of farms. Some interesting experiments are underway where farmers are visited 4-5 times per year by an inter-

viewer trained to collect data on events in the intervening period. Such efforts should eventually show how the high costs of data collection can be reduced, while maintaining reasonable control over the quality of the information.

IV. Data Collection and Analysis Problems

In this section I would like to elaborate further on some of the problems frequently encountered in collecting and analyzing data from cross-sectional surveys. The first problem is that of the instruments used in collecting data. Many questionnaires are so long, complicated, and unwieldy that the interview takes much more time than can be considered ideal for good quality responses. Why does this so often happen even when the researchers understand preferred interviewing techniques? One reason is that the objectives of the research or the specific data requirements are not clear. In such cases the reaction becomes, "Let's ask the question just in case we might want the information later." Furthermore, to keep down costs, data needs for one project are "piggy-backed" on the survey for another project and usually both pieces of research suffer. A sharp role distinction between researcher and data collector can also contribute to the problem. The feasibility of the survey may be ignored if the designer of the research sends the unsupervised survey crew to collect data with his "ideal" questionnaire, while he awaits their return reading journals or politicing with the minister in the capital.

Large surveys create a serious problem of quality control of information. In our surveys, we found it necessary to have one fulltime supervisor checking questionnaires for every five interviewers each of which completed 1-2 interviews per day. Each questionnaire was completely

checked for missing, incomplete or inconsistent data within a day of the interview, while the survey crew was still in the area, so the interviewer could return to the farm if necessary. For a large survey conducted under strict time constraints, good supervisors may not be available in the large numbers required, or most of the resources for the research may be spent on interviewers, drivers and vehicles. The result can only be the subsequent elimination of a large number of questionnaires with faulty data and uncertain quality of many of those retained.

When plans and budgets are made for research, all too frequently the importance of the data processing and analysis phase is minimized. Most of the resources are spent on survey costs and researcher salaries with little left for coding, keypunching, data editing, computer time and programming. The researchers plan their schedules assuming the data will be available soon after the survey and, in turn, the sponsors and policymakers expect the results soon afterwards. The long unexpected delays which usually occur due to inadequate planning and funding contribute to policymakers' frustration over the capability of research to resolve their immediate needs.

This phase of research is obviously related to the complexity of the questionnaire and number of observations. A survey with several hundred observations, 2-3,000 variables and a non-precoded questionnaire can easily require 3-4 months of time for preparation and duplication of the code, coding the data, punching and verifying the cards. Computer consistency checks, identification of errors and questionable data and verification with the original questionnaires can require several additional weeks unless the entire procedure including writing of computer programs is developed before or simultaneously with the survey.

Qualified computer programmers frequently present a bottleneck. Staff with knowledge of the region being studied are also required to go back into the original questionnaires and resolve problems identified in the computer checks. A large survey with a long questionnaire creates problems for developing simple programming packages and for institutions with limited computer capacity.

Much of the data processing and analysis represents an uneven work load with occasional peak programming and computer needs followed by lulls with limited activity. Thus most computing facilities need to work on several projects simultaneously, each with its own set of needs and associated time schedule. Often peak needs occur at the same time and researchers must wait their turn and schedules may get delayed even more.

In a large survey where several research projects are involved, it may be difficult to get all of the data cleaned, checked and stored at one time. Each researcher with his own time schedule faced with these delays is tempted to work only with that data essential for his immediate needs. Rather than take the necessary time to carefully check questionable values, he may be tempted to assume that errors are compensating or simply throw out uncertain observations.

Two obvious solutions are possible in the face of these processing and analysis problems. First, both researchers and policymakers need to recognize the great amount of time and resources required for this stage of research, and plan and staff accordingly. Alternatively we can drop the great emphasis on huge sample sizes and concentrate on doing more work more carefully and quickly with a manageable amount of data. My preference is to run the risk of limited representativeness in order to quickly identify problems in a changing economy rather than carefully

plan a definitive piece of research requiring massive amounts of data thereby delaying the results until such time as the problem is abundantly clear to everyone without the need for the research.

IV. Conclusion

The basic argument of this paper can be summarized as an appeal 1) to identify who should do what research on financial markets in any given LIC, 2) to use existing data bases for research before rushing out to do more poorly planned massive surveys which consume limited research resources and talent and 3) to recognize the magnitude of the problems and time involved in processing and analyzing mountains of data. As economists, we are accustomed to dealing with allocation of resources and production efficiency. These same concepts need to be applied to research endeavors. We also need to help the policymakers, as the chief users of research, understand the problems of conducting research and the limitations that researchers have in effectively meeting their short run policy needs. This workshop should provide one more opportunity for identifying the proper role of research and researchers in the development of rural financial markets.